

**CSE 2194: Supervised Machine Learning**  
**Programming Assignment-III**  
**(Data Visualization: part-2)**

**Question 1:**

Download the dataset given in the following link in your local repository. Read the dataset and clean it if it has some missing values, wrong data, wrong formats or duplicate values. Clean the dataset and print the 5 to 15 rows of the data. Finally, save the clean dataset in your local repository.

Dataset: [data.csv](#)

**Question 2:**

Given two arrays, `arr_1 = np.array([1, 2, 3, 4])`, `arr_2 = np.array([2, 4, 6, 8])`. Print the results of the following operations: Add, Subtract, Multiply, Divide. Use NumPy

- A. Generate a random integer 2D array with three rows and four columns named `arr_3`, with four values between 0 and 100.
- B. print the exponential and logarithmic values of all elements in array `arr_1`.
- C. Given an array `arr_4 = np.array(1.2, 2.5, 5.6, 3.4, 7.8]`, print the round-down (floor) and round-up (ceil) values.

**Question 3:**

Generate an array of 200 values between  $-\pi$  &  $\pi$ . Calculate the corresponding sin, cos and tan value for the generated array. Finally, plot the sin, cos and tan curves using the matplotlib library.

**Hints:** use `np.linspace` to generate values (`gen_arr`), then `sin_values = np.sin(gen_arr)` and so on for others and finally use `plot(gen_arr, sin_values)`. Follow the same for cos and tan also.

**Question 4:**

Given two 2D arrays, `arr_1 = np.array([[1, 2], [3, 4]])`, `arr_2 = np.array([[2, 4], [6, 9]])`.  
Perform the following operations.

- A. Matrix multiplication with Dot Product
- B. Compute eigenvalues and eigenvectors for both matrices.
- C. Compute the determinant of both matrices.
- D. Compute the inverse of both the matrices.

Hint: Use `from numpy import linalg as LA`

### Question 5:

Using the given data, use the matplotlib library to draw the Pie chart for the column class.

Dataset: [Data](#)

### Question 6:

A pairwise plot is a favourite in exploratory analysis to understand the relationship between all possible pairs of numeric variables. Use the seaborn library to load the iris dataset and then plot the pairwise plot for the iris data.

### Question 7:

Download the given dataset and plot the histogram for integer and float columns of the dataset.

Dataset: [Data](#)

### Question 8:

Download the dataset and plot the scatter plot for the sale price and profit columns.

Dataset: [Data](#)

### Question 9:

Download the given dataset and split the dataset into 70% for training and 30% for testing using the Scikit library.

Dataset: [Data](#)

### Question 10:

Download the dataset given in the link, convert the categorical data into integer columns, and print the shape and head of the dataset.

Dataset: [Data](#)

## Home Assignment:

### Question 1:

Read the dataset given in the following link. Print the info on the data and clean the data if required. Visualize the data using an appropriate diagram based on your observations.

Dataset: [Data](#)

### Question 2:

The time series plot visualises how a given metric changes over time. Use the given data and draw the time series plot to visualise how the Air Passenger traffic changed between 1949 and 1969.

Dataset: [Data](#)

### Question 3:

Read the dataset given in the following link. Print the info on the data and clean the data if required. Finally, draw the bar chart, where the x-axis has the ['manufacturer'] column, and the y-axis has the ['counts'] column.

Dataset: [Data](#)

### Question 4:

A correlogram is used to visually see the correlation metric between all possible pairs of numeric variables in a given data frame (or 2D array). Import the dataset provided in the following link and plot the heatmap to visualize the correlation.

Dataset: [Data](#)

-----\*\*\*\*\*-----